



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/843,255	04/24/2001	Ramarathnam Venkatesan	MS1-647US	7357
22801	7590	07/27/2007	EXAMINER	
LEE & HAYES PLLC			BALI, VIKKRAM	
421 W RIVERSIDE AVENUE SUITE 500				
SPOKANE, WA 99201			ART UNIT	PAPER NUMBER
			2624	
			MAIL DATE	DELIVERY MODE
			07/27/2007	PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>
	09/843,255	VENKATESAN ET AL.
	<b>Examiner</b>	<b>Art Unit</b>
	Vikkram Bali	2624

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --  
**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### **Status**

- 1) Responsive to communication(s) filed on 17 May 2007.
- 2a) This action is **FINAL**.                            2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### **Disposition of Claims**

- 4) Claim(s) 1-7, 15-19, 64, 65 and 67-75 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) Claim(s) \_\_\_\_\_ is/are allowed.
- 6) Claim(s) 1-7, 15-19, 64, 65 and 67-71 is/are rejected.
- 7) Claim(s) 72-75 is/are objected to.
- 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### **Application Papers**

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on \_\_\_\_\_ is/are: a) accepted or b) objected to by the Examiner.  
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### **Priority under 35 U.S.C. § 119**

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All    b) Some \* c) None of:
  1. Certified copies of the priority documents have been received.
  2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

#### **Attachment(s)**

- 1) Notice of References Cited (PTO-892)
- 2) Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) Information Disclosure Statement(s) (PTO/SB/08)  
 Paper No(s)/Mail Date 11/17/06; 2/28/07; 5/17/07.
- 4) Interview Summary (PTO-413)  
 Paper No(s)/Mail Date. \_\_\_\_\_
- 5) Notice of Informal Patent Application
- 6) Other: \_\_\_\_\_

**DETAILED ACTION**

***Claim Rejections - 35 USC § 103***

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

3. Claims 1-7, 15-19, 64-65 and 67-71 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hull et al (US 5465353) in view of Pugh et al (US 6658423).

Regarding Claim 1, Hull discloses a computer-implemented method for hashing a body of text, the method comprising:

obtaining a body of text containing textual content in a computer readable format (Figure 2, numerical 202, the textual content on the paper as to be scanned using the

scanner 204), and wherein the textual content of the obtained computer readable formatted body of text is mutable via software tools for manipulation of textual content of bodies of text, (reference Hull discloses scanner (figure 2, numerical 204), and the page (fig. 2, numerical 202), once it [page] is scan one [individual] can easily manipulate the scanned document via a software such as MS-Word or like “mutable via software tools”);

formatting the filtered subtext into a defined image based format, wherein the textual content of the formatted filtered subtext is immutable via software tools for manipulation of the textual content, (the scanner convert “format” the textual data on the paper numerical 202 of the figure 2 into an image, and Hull in the figure 2 once the scanner 204 scans the document 202 to create an image then the textual content of the document cannot be manipulated by the software tools, because the textual content requires different software tools to manipulate then the image in the computer system);

deriving a hash value representative of content of the filtered subtext, perceptually distinct filtered subtexts having hash values that are substantially independent of each other, **the deriving comprises hashing the image based formatted filtered subtext resulting from the formatting** (Figures 2, numerical 214, and see col. 13 lines 30-44, wherein the hash descriptors are created using the word of the document and the descriptors are “substantially independent of each other” as they are dependent on the lengths of the words).

However, he fails to explicitly disclose filtering the textual content the body of the text to remove elements of the textual contents, wherein filtering act products filtered subtext, which is a subset of the textual content of the body of text, as claimed.

Pugh in detecting duplicate and near duplicate files teaches filtering the textual content the body of the text to remove elements of the textual contents, wherein filtering act products filtered subtext, which is a subset of the textual content of the body of text, (col. 7, lines 45-55, in that specifically line 50-51 states extracting parts from the documents i.e. "filtering" and the words is read as the "subtext") as claimed.

It would have been obvious to one ordinary skilled in the art at the time of invention to combine the two references as they are analogous because they are solving the similar problem of document matching. Incorporating the teachings of Pugh of filtering the document in order to extract the words can be incorporated in the Hull system in order to attain a system that provides a better near duplicate determination technique, (see col. 7, lines 26-27) for motivation.

Regarding Claim 2, Hull further discloses a method as recited in Claim 1, wherein perceptually distinct image based formatted, filtered subtext have hash values that are independent of each other (col. 13 lines 30-44, wherein the hash descriptors are created using the word of the document and the descriptors are different as they are dependent on the lengths of the words "perceptually distinct bodies of text have hash values").

Regarding Claim 3, Hull further discloses a method as recited in Claim 1 further comprising comparing hash values of two image based formatted filtered subtext to

determine if such values match (col. 9, lines 35-40, wherein the hash descriptors are cross referenced in order to query the document database “comparing hash values of two bodies of text to determine if such values match”).

Regarding Claim 4, Hull further discloses a method as recited in Claim 1 further comprising comparing hash values of two image based formatted filtered subtext to determine if such values substantially match (col. 9, lines 35-40, wherein the hash descriptors are cross referenced in order to query the document database “comparing hash values of two bodies of text to determine if such values substantially match”).

Regarding Claim 5, Hull further discloses a method as recited in Claim 4 further comprising indicating whether such values substantially match (col. 9, lines 35-40, wherein the hash descriptors are cross referenced in order to query the document database and figure 3, the accumulator sets all the documents that has the same hashed descriptors “indicating whether such values substantially match”).

With regards to Claims 6, 7 and 64, arguments analogous to those presented for Claim 1 are applicable to Claims 6, 7 and 64.

With regards to Claims 15-17, arguments analogous to those presented for Claims 1-5 are applicable to Claims 15-17. Hull further discloses similar bodies of text having proximally similar hash values (Figures 2, numerical 214, and see col. 13 lines 30-44, wherein the hash descriptors are created using the word of the document and the descriptors are “proximally similar hash values” as they are dependent on the lengths of the words, and if the words are same the hash values will be proximally be similar) as claimed.

With regards to Claims 18, 19 and 65, arguments analogous to those presented for Claims 15-17 are applicable to Claims 18, 19 and 65.

With regards to the claim 67, Pugh further teaches indicating suspicion of plagiarism between the two filtered subtext when the compared hash values of the two filtered subtext substantially match, (see col. 1, lines 15-21) as claimed.

With regards to the claims 68 and 70, Hull further teaches the text comprises multiple words and sentences, (see col. 13, lines 20-30 the two documents use for the example for creating the hash) as claimed.

With respect to claim 69 and 71, Hull teaches the test comprises multiple words a sentences, and hash is representative of all the textual content (see col. 13, lines 20-30 the two documents use for the example for creating the hash, and col. 13 lines 30-44, wherein the hash descriptors are created using the word of the document and the descriptors are dependent on the lengths of the words in the document) as claimed.

#### ***Allowable Subject Matter***

4. Claims 72-75 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

#### ***Response to Arguments***

5. Applicant's arguments filed 5/17/2007 have been fully considered but they are not persuasive. Applicant argues that none of the references discloses the newly added

limitation of "hashing the image based formatted, filtered subtext resulting from the formatting" as claimed (see remarks [0020] page 16) of 5/17/2007 arguments.

Examiner disagrees and would like to point out that the claims are given their broadest reasonable interpretations. And, for the rejection please see the claim rejection above.

### ***Conclusion***

6. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Vikkram Bali whose telephone number is 571.272.7415. The examiner can normally be reached on 7:00 AM - 3:30 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Eileen Lillis can be reached on 571.272.6928. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Vikkram Bali  
Primary Examiner  
Art Unit 2624

vb  
July 23, 2007